caBIG™
cancer Biomedical Informatics Grid ™

The caBIG™ Data Sharing and Security Framework:

Introduction and Demonstration

Wendy Patterson, J.D.
NCI Facilitator, DSIC Workspace
Senior Advisor, NCI Technology Transfer Center

Marsha Young, J.D.
DSIC Workspace Lead
Associate, Booz Allen Hamilton

Rachel Nosowsky, J.D.
DSIC Workspace Member
Assistant General Counsel, University of Michigan

January 2008

Outline for Today

- Present the caBIG™ Data Sharing and Security Framework ("DSSF" or "Framework")
- Review currently available tools available to assist researchers and institutions in implementing the Framework
  - NCI caGrid Host Trust Agreement for Non-Sensitive Data
  - Researcher Questionnaire
    (For assessing data sharing restrictions)
  - Guidelines for Preparing Data Sharing Plans
    (For review by IRB and/or other institutional officials)
- Preview select tools in development

Why Share Data?

- Increasingly, the large volumes of research data created by high throughput genomics and proteomics technologies can best be harvested by teams of individuals - rather than by single PI's.
- To realize the scientific and public health benefits of translational and personalized medicine, collaboration across and within disciplines is required to leverage broad knowledge and skill bases.
- Data sharing raises the visibility of individual studies and data collections; it opens avenues of data dissemination and validation – leading to more citations in publications and increased prominence.

caBIG™
cancer Biomedical Informatics Grid ™

Understanding the caBIG™ Data Sharing and Security Framework

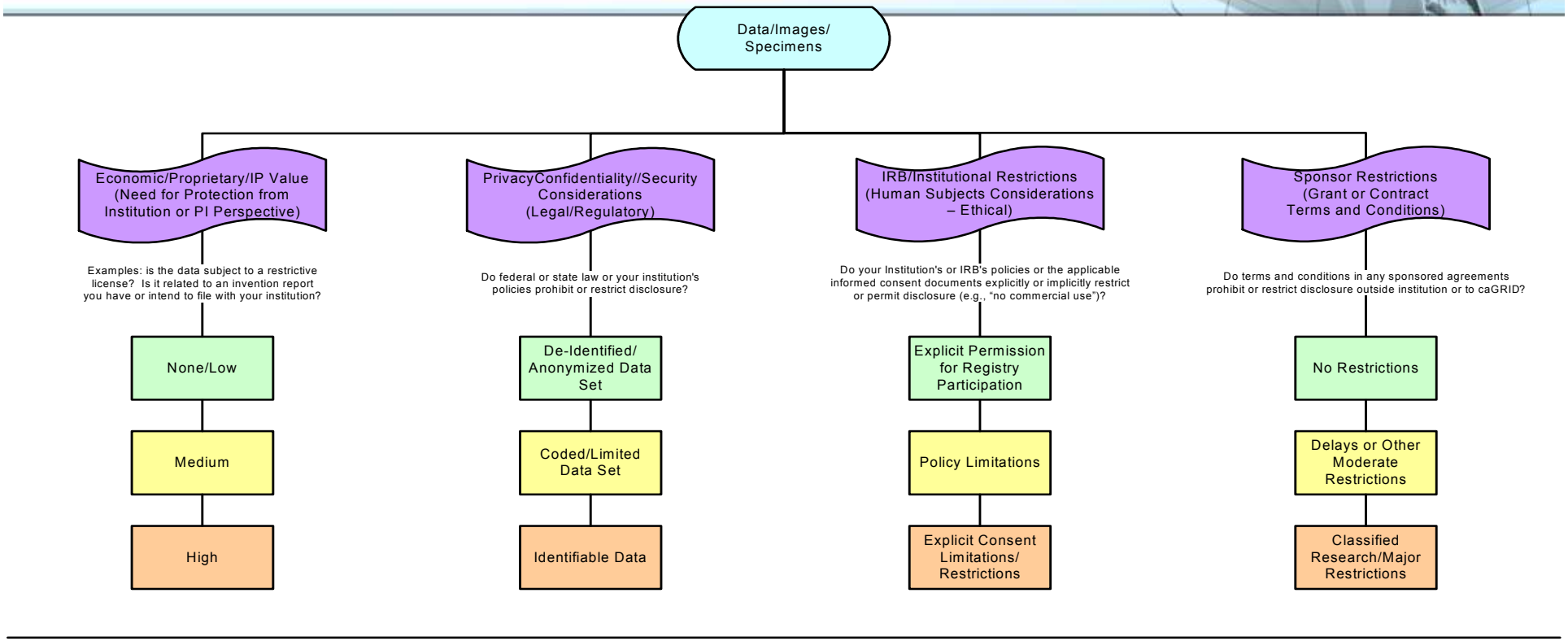
The Security Elements of the Framework

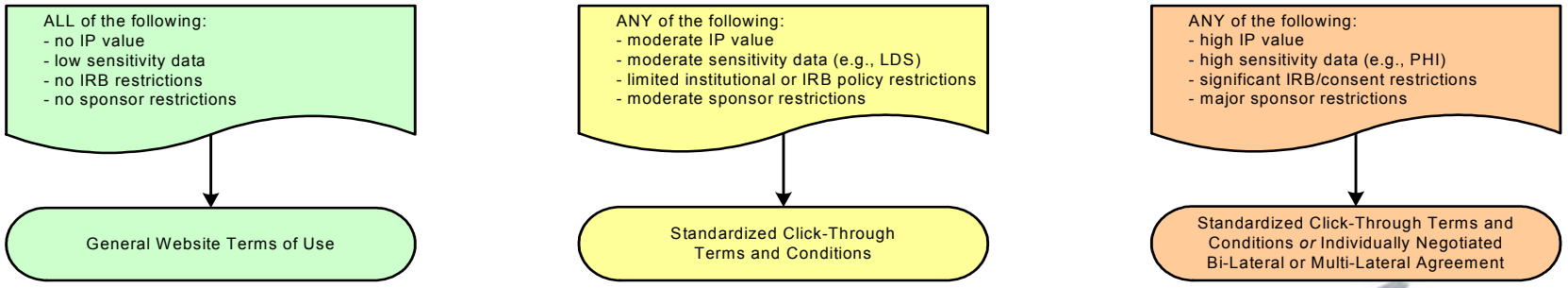
- The caBIG™ policies and procedures that encompass the security and data access controls – *authentication and authorization* -- that enable Centers to share data in compliance with federal and state requirements and local institutional policies.

- The requirements for authentication of identity associated with low, medium and high sensitivity datasets are based on National Institute of Standards and Technology (NIST) guidance for authentication at known Levels of Assurance (LOA) and are informed by use cases from the community.

- These Levels of Assurance reflect the extent of "proof" that the users are indeed who they assert they are, e.g. LOA 1 requires no proof or assurance. LOA 2 requires some presentation of a government-issued ID for inspection by a credential provider, and so forth.

caBIG™ Data Sharing and Security Framework





Understanding the Framework

- **Purpose of the Framework**
  - Determine which data can be shared
  - Identify necessary access and data security controls (authentication, authorization)
- **Audiences**
  - IRBs
  - Privacy Officials
  - Industry-Sponsored Projects/Grants & Contracts Administration/Tech Transfer Officers
  - Institutional Attorneys
- **Function (Summary)**
  - Assess data sensitivity by reference to the Framework's four principal elements:
    - **Privacy Considerations**
    - **IRB/Ethical Restrictions**
    - **Proprietary Value (to Researcher/Institution)**
    - **Sponsor Requirements**
  - Assign a low, medium or high sensitivity rating to the data
  - Use sensitivity assignment process to determine what data will be shared and the appropriate mechanism for sharing

Understanding the Framework

- **Sensitivity assessment**
  - Data sensitivity is judged – by the organization that owns, controls, or manages the data (the "Provider") – first along each of the four axes referenced above
  - Sensitivity is influenced by many factors, including
    - The organization's interpretation of governing laws, regulations, and ethical standards (federal, state, local)
    - Organizational policies
    - Contractual constraints
  - The outcome, that is, a low, medium or high sensitivity rating, determines what data will be shared, when, and how
- **The Provider determines necessary access controls by assessing:**
  - The level of certainty needed to *authenticate* data users ("Recipients")
  - Whether particular authenticated groups or individuals are *authorized* to access the particular data

Using the Framework as an Analysis Tool

- **Providers can use the Framework to help determine the structures and mechanisms needed to share the data under consideration:**

General Website Terms of Use -
Minimum Restrictions on Access

Standardized Click-Through Terms and Conditions -
Some Limitations on Access to the Data

More Restricted Access Conditions -
(Could be Standardized or Individually Negotiated Agreement)

Note: **Each Provider must select the type of data sharing mechanism that best fits its needs. The Data Sharing and Security Framework is not a strict policy or guideline, but instead a means of analysis.**

caBIG

Data Sharing in the Green Lane

Minimum Restrictions on Access

- Provider/Grid Host (individual or institutional) must:
  - No authorization conditions on users and therefore no contract (MTA, DUA, etc.) required
  - Enter into the LOA1 NCI-caGrid Host Agreement – agree to
    - *Only make non-sensitive data available through NCI-caGrid*
    - *Operate in accord with fairly common security practices -- apply patches for known/published "bugs" or prevent known security attacks*
- Recipient/Grid User must:
  - Accept General Website Terms of Use
  - Authenticate to LOA1 or LOA2 of the NIST Guidelines

Data Sharing in the Yellow Lane

Standardized Click-Through Terms and Conditions -

Some Limitations on Access to the Data

- Provider/Grid Host must:
  - Enter into a trust agreement – *in development*
  - Adopt model caBIG™ standardized click-through agreement that specifies some restrictions on data sharing – *in development*
- Recipient/Grid User must:
  - Agree to terms of standardized click-through agreement
  - Enter into an agreement with NCI or 3rd party authentication certificate provider concerning authentication (NIST LOA 2 or 3) – *in development*

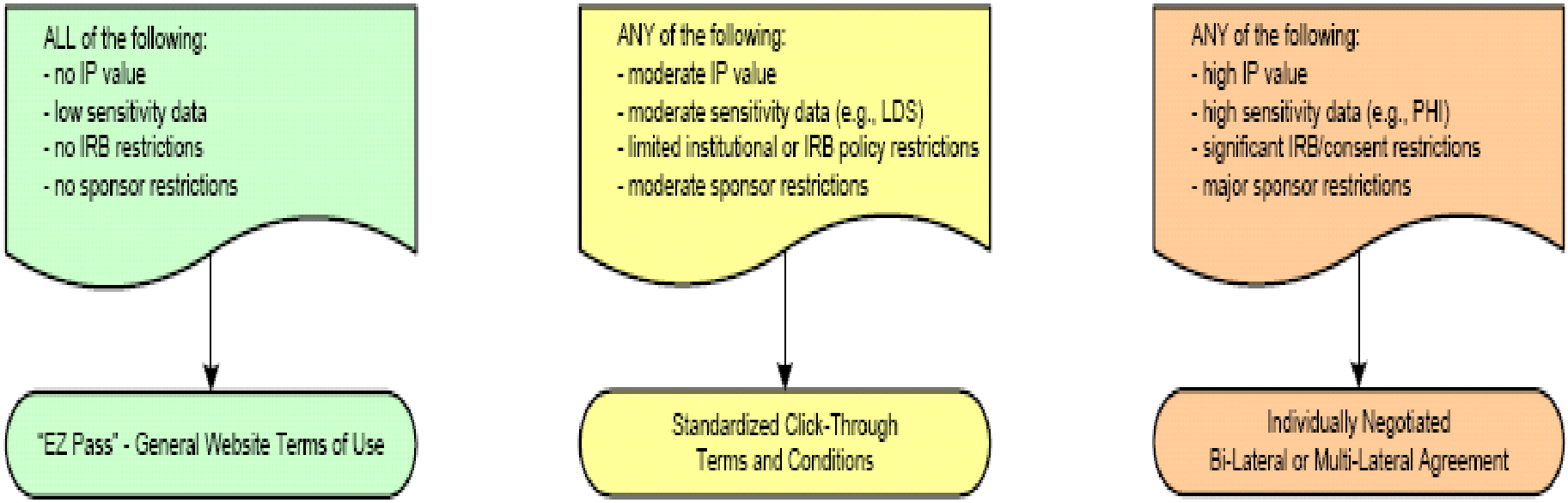
Data Sharing in the Orange Lane

More Restricted Access Conditions - Could be Standardized or Individually Negotiated Agreement

- Provider/Grid Host must:
  - Enter into a trust agreement – *in development*
  - Specify restrictions on data sharing in model caBIG™ standardized terms and conditions OR adapt institutional agreement to adhere to caBIG™ standards/guidelines for negotiating such agreements – *BOTH in development*
- Recipient/Grid User must:
  - Agree to Provider/Grid Host's terms and conditions (adopted or adapted)
  - Enter into an agreement with NCI or 3rd party authentication certificate provider concerning authentication (NIST LOA 3 or 4) – *in development*

Assessing Data Sensitivity

After conducting the sensitivity assessment, the providing researcher/institution can then select an appropriate data sharing mechanism.



The assessment process is discussed on the following slides

Assessing Data Sensitivity – Privacy Considerations

- PRIVACY: **Do federal or state laws, or your institution's privacy or confidentiality policies, restrict disclosure (research or IRB policies are addressed separately)?**
  - ***The Framework asks Providers/Grid Hosts to select the category of sensitivity that best describes their data:***

Low: De-Identified/Anonymized Human Data Set or Non-Human Data

Medium: Coded/Limited Data Set

High: Individually Identifiable Data

caBIG

Assessing Data Sensitivity – IRB/Ethical Restrictions

- Relates to Human Subjects Research Considerations: **Do the Common Rule, FDA regulations, or your institutional or IRB policies, or the applicable informed consent documents, explicitly or implicitly restrict or permit disclosure (e.g., limitations on use for new or unrelated projects, prohibition on commercial use, etc.)?**

- *The Framework asks Providers/Grid Hosts to select the level of restriction that best describes their data:*

Low: Explicit Permission for Registry Participation

Medium: Policy or Consent Limitations

High: Explicit Consent Limitations/Restrictions

caBIG

Assessing Data Sensitivity – Proprietary Value

- Relates to the Data's Value to the Organization or Researchers: **Do the data relate to an invention report a researcher has, or intends to file, with his/her institution? Are the data or study findings awaiting publication?**
- *The Framework asks Providers/Grid Hosts to select the category of proprietary value that best describes their data:*

None/Low: Not subject to restrictive license or invention report

Medium: Data not yet submitted for publication

High: nonpublic intellectual property or other significant restriction

Assessing Data Sensitivity – Sponsor Requirements

- Relates to Requirements of Sponsors in Grants and Contracts: **Do the terms and conditions in any sponsored agreements prohibit or restrict disclosure outside the institution or via the Grid? Are the data subject to a restrictive license?**

- ***The Framework asks Providers/Grid Hosts to select the level of sponsor restriction that best describes their data:***

Low: No Restrictions

Medium: Delays or Other Moderate Restrictions

High: Classified Research/ Major Restrictions

caBIG™
cancer Biomedical Informatics Grid ™

Implementing the caBIG™ Data Sharing and Security Framework:

Working with the Tools

Implementation Tools

- **Tools ready for piloting:**
  - The Framework (already discussed)
  - Level 1 NCI-caGrid Host Agreement
  - Researcher Questionnaire (for assessing data sharing restrictions)
  - Guidelines for Preparing Data Sharing Plans (for review by IRB or other institutional officials)
- **Tools next up for testing:**
  - Model Informed Consent
  - Security Policies for Moderately Sensitive Data
- **Tools in development:**
  - Guidelines for de-identifying data
  - Standardized data use agreements/medium sensitivity data
  - Standards for developing agreements/high sensitivity data
  - Data access authorization policies for medium to high sensitivity data

Guidelines for Preparing Data Sharing Plans

- **Purpose**
  - Facilitates identification of data to be shared and mechanism for sharing
  - Provides a framework for organizing information important to various data stewards or other interested institutional officials
- **Background information related to the institution and project where caBIG™ tools will be adopted**
  - Identifies issues that will drive legal/regulatory determinations related to data sensitivity, need for IRB oversight, etc.
  - Identifies purpose/objectives of adoption project
  - Identifies who may have an interest in the data – proprietary, regulatory, etc.
- **Information about the data to be shared**
  - Sources
  - Summary of data elements
  - Intended recipients
  - Mechanisms for data sharing (access controls, etc.)
  - Timing
- **Information about institutional units (including IRBs) who must approve data sharing plans**
- **Open-ended question regarding additional anticipated challenges**

NCI/caGrid Host Agreement – Level 1 Data

- **Required for hosting Level 1 data services; intended for sharing non-sensitive data (de-identified or non-human)**
- **Can be signed by individuals who host Grid services; they determine if additional review/signature is required**
- **Purpose: describe information security responsibilities of Grid Hosts**
- **Grid Host is responsible for:**
  - Complying with applicable laws and regulations
  - Implementing policies and procedures to enable compliance with caBIG™ security principles
  - Reporting security breaches and participating in security investigations
  - Applying system upgrades and patches
  - Maintaining security of host certificates

Informed Consent

- **Model language re: caBIG™ for use in pre-existing authorization/consents**
  - Provide basic language for use by adopters (and others who want to facilitate data sharing) within their own documents; facilitate adoption of caBIG™ language in other models under development by various other groups

- **Standardized choices for research participants related to specimen use and/or data sharing**
  - Standardize choices in authorization/consent forms using process similar to other caBIG™ standards development activities; facilitate adherence to patient/participant choices

- **Model informed consent and HIPAA authorization document – "the whole package" (disclosure + options)**
  - Assist institutions and smaller provider-based participants in drafting informed consent/authorization forms compliant with Common Rule, FDA and HIPAA requirements that will facilitate data sharing across the Grid

  - ***STATUS:*** *Preliminary drafts complete; revised drafts to be posted during January 2008; additional refinements expected following January DSIC F2F and solicitation of comments/feedback from adopters, developers, and others (including VCDE); final approval expected in late winter/early spring 2008*

Security Policies and Procedures for Low to Medium Sensitivity Data

- Governance Model for Authentication for services available to persons who authenticate at LOA 2
  - caBIG™ Identity Provider Federation Policy document (to describe governance model)
  - Model Trust Agreement for Interfederation (to bring new identity providers into the trust federation)
  - LOA 2 Technical Implementation (based on understanding of policies and issues related to implementation identified by NCI-caGrid Security Working Group
  - ***STATUS:*** ***work in progress with SWG; deliverables expected late winter/early spring***

Framework Tools in Development

- **DSIC Workspace members are working on developing additional elements of the DSSF Bundle, such as:**
  - Guidelines for de-identifying data- the “whys” and “hows” of data de-identification practices to comply with HIPAA regulations
  - Standardized, click-through data use agreement for use with medium sensitivity data; standards for developing agreements for sharing high sensitivity data
  - Data access authorization policies for medium to high sensitivity data for institutional groups and unaffiliated users; model trust agreement for authorization

- **“Getting Connected” Centers are expected to implement the Framework, along with using the tools supporting the Framework through testing them within the institution’s workflow, as appropriate, and providing feedback to the DSIC Workspace.**

- **Centers may also assist in developing, testing and refining DSSF Bundle elements through active participation of institutional subject matter experts in DSIC Workspace activities.**

Next Steps

- **If you are in a Cancer Center participating in the "Getting Connected with caBIG™" deployment effort, you need to complete the following next steps to implement the caBIG™ Data Sharing and Security Framework:**

1. Visit the DSIC Workspace to review workspace activities as you identify qualified Cancer Center representative (s) who can best contribute to the review and development of workspace products, such as model documents, policies and best practices.
2. Contact the DSIC WS Lead, Marsha Young (young_marsha@bah.com) with your representative's name, contact information, and area of interest/expertise.
3. Join the DSIC listservs to receive announcements and documents.
   CABIG_DSIC-L - Data Sharing and Intellectual Capital Workspace
   CABIG_DSIC_PRO_SIG-L - Working Group - Proprietary SIG
   CABIG_DSIC_REG_SIG-L - Working Group - Regulatory SIG

Questions

Questions???